One of the key challenges that researchers of multimodal communication face, is that the empirical analysis of speech in relation to gestural behavior, gaze and other modalities requires high-quality video data and detailed annotation of the different semiotic resources under scrutiny. In the majority of cases, the annotation of hand position, hand motion, gesture type, etc. is done manually, which is a time-consuming enterprise requiring multiple annotators. In this paper we present a semi-automatic alternative, in which the focus lies on minimizing the manual workload while guaranteeing highly accurate gesture annotations. More specifically, we zoom in on three features of our system: (i) segmentation-based hand detection, (ii) positioning of hands in gesture space, and (iii) analysis of the directionality of gestures.

Our gesture analysis builds on a semi-automatic, segmentation-based, hand detection approach as proposed by [3]. Once the positions of the hands are obtained, our framework automatically segments a recording in gesture and non-gesture segments based on the position of the hands. We validated our gesture segmentation on a recording of the NeuroPeirce corpus [1] and of the SaGA corpus [4], since these data sets include manual annotations of gesture segments. In total, both recordings have a duration 13.5 minutes and consist of a total of 23500 video frames. Comparing our automatic gesture segmentation against manual segmentation resulted in an average F1-accuracy of 88.61%, which demonstrates the usefulness of our approach. Furthermore, the manual effort is reduced to a minimum: in only 2.6% of the frames, the system required manual validation or correction.

In a second step, the result of the gesture segmentation described above is used as a basis for calculating the position of the hands in gesture space. Manually annotating the gesture space is extremely labor-intensive, since ideally, one has to assign a specific spatial position to each individual frame of a gesture sequence. For that reason, most manual annotations of spatial information only provide one value for an entire gesture phase or unit. To overcome this problem, our approach automatically analyzes the position in gesture space for each gesture segment according to McNeill's gesture space [5] and automatically defines the appropriate sector and sub-sector for each hand in each frame of the segment.

A third analytical layer concerns the directionality of gestures. Several gesture annotation systems (e.g. [2]) and empirical accounts include the direction and movement of hand gestures, resulting in a specific trajectory. Comparable to the positioning in gesture space (cf. ii), we noticed that manual annotation is often restricted to a partial analysis. For example, the directionality of an entire leftward pointing gesture is often annotated as "left", since this is the major direction of movement. To further support and refine the annotation, we propose an automatic alternative. Here, we calculate the direction of movement for each frame by comparing the hand positions of the current frame and the positions in the previous frame. This generates a reliable and fine-grained movement analysis that can be used for further statistical and time-sensitive analysis.

References

[1] Brenger, B., and Mittelberg, I. Shakes, nods and tilts. motion-capture data profiles of speakers' and listeners' head gestures. In Proceedings of the 3rd Gesture and Speech in Interaction (GESPIN) Conference (2015), pp. 43–48.

[2] Bressem, J. Transcription systems for gestures, speech, prosody, postures, and gaze. In Proceedings of Body - Language - Communication: An International Handbook on Multimodality in Human Interactiong (2013), vol. 1, pp. 1037–1059.

[3]De Beugher, S., Brône, G., and Goedemé, T. Semi-automatic hand annotation making human-human interaction analysis fast and accurate. In Proceedings of the 11th International Conference on Computer Vision Theory And Applications (VISAPP) (Rome,Italy, 2016), pp. 552–559.

[4]Lücking, A., Bergmann, K., Hahn, F., Kopp, S., and Rieser, H. The Bielefeld Speech and Gesture Alignment Corpus (SaGA). In Proceedings of LREC 2010 Workshop: Multimodal Corpora–Advances in Capturing, Coding and Analyzing Multimodality (2010), pp. 92–98.

[5]McNeill, D. Hand and Mind: What gestures reveal about thought. University of Chicago Press, Chicago, Illinois, 1992.